HerosNet: Hyperspectral Explicable Reconstruction and Optimal Sampling Deep Network for Snapshot Compressive Imaging

Xuanyu Zhang[1], Yongbing Zhang[2], Ruiqin Xiong[3], Qilin Sun[4], Jian Zhang[1,5]

[1]Peking University Shenzhen Graduate School, Shenzhen, China
[2]Harbin Institute of Technology, Shenzhen, China   [3]Peking University, Beijing, China
[4]The Chinese University of Hong Kong, Shenzhen, China   [5]Peng Cheng Laboratory, Shenzhen, China

Abstract

*Hyperspectral imaging is an essential imaging modality for a wide range of applications, especially in remote sensing, agriculture, and medicine. Inspired by existing hyperspectral cameras that are either slow, expensive, or bulky, reconstructing hyperspectral images (HSIs) from a low-budget snapshot measurement has drawn wide attention. By mapping a truncated numerical optimization algorithm into a network with a fixed number of phases, recent deep unfolding networks (DUNs) for spectral snapshot compressive sensing (SCI) have achieved remarkable success. However, DUNs are far from reaching the scope of industrial applications limited by the lack of cross-phase feature interaction and adaptive parameter adjustment. In this paper, we propose a novel **H**yperspectral **E**xplicable **R**econstruction and **O**ptimal **S**ampling deep **Net**work for SCI, dubbed **HerosNet**, which includes several phases under the ISTA-unfolding framework. Each phase can flexibly simulate the sensing matrix and contextually adjust the step size in the gradient descent step, and hierarchically fuse and interact the hidden states of previous phases to effectively recover current HSI frames in the proximal mapping step. Simultaneously, a hardware-friendly optimal binary mask is learned end-to-end to further improve the reconstruction performance. Finally, our HerosNet is validated to outperform the state-of-the-art methods on both simulation and real datasets by large margins. The source code is available at https://github.com/jianzhangcs/HerosNet.*

1. Introduction

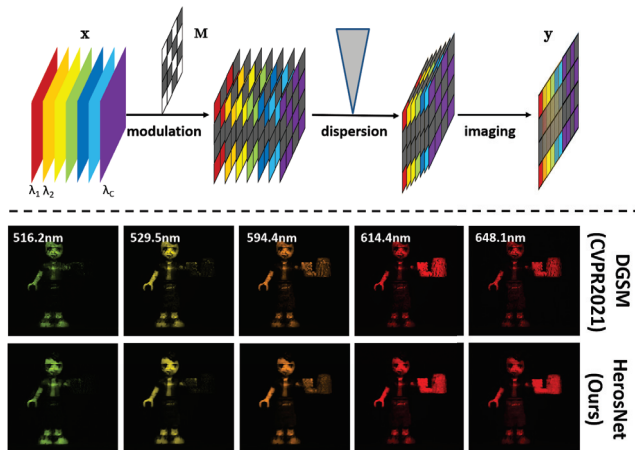With the development of artificial intelligence and robotics, the demand for capturing and sensing hyper-

Figure 1. The schematic of CASSI system and some visual results of the proposed HerosNet and DGSM [13] on the real dataset. Our reconstructed HSIs have clearer edges and more detailed textures, while the results of DGSM have more noise and artifacts.

spectral images has dramatically increased in recent years [3, 6, 12, 47]. Based on the traditional compressive sensing (CS) [51, 54], spectral snapshot compressive sensing (SCI) system aims to record 3D scenes via a 2D detector. It has the advantages of low bandwidth, low cost, and high data throughput, which has played an increasingly pivotal role in a wide range of applications, such as remote sensing, object detection, super-resolution, and medical diagnosis [2, 8, 14, 18, 25, 36, 38, 55]. In this paper, we focus on a typical imaging system named coded aperture snapshot spectral imager (CASSI) [11, 21, 30], which modulates spectral frames via a coded aperture (i.e. physical mask) and shifts them across the spectral dimension via a disperser.

In the past few years, there have been a great amount of reconstruction methods for HSIs compressed by CASSI, including model-based methods and deep learning-based methods. Traditional model-based methods [5, 15, 16, 29, 34, 40, 41, 46] often search for the optimal solution it-

eratively and refine the results to the desired signal domain via the image priors. Although these methods are highly interpretable, they are limited by the hand-crafted priors and slow reconstruction speed. Owing to the development of deep learning, several learning-based methods [20, 22–24, 26, 39, 43, 56, 57] have been used to reconstruct HSIs, which directly learn an end-to-end inverse mapping from the 2D measurement to the 3D hyperspectral cube. Compared to model-based methods, they tend to drastically reduce time complexity and achieve better performance. However, they do not explicitly embody the system imaging model and are just trained as a black box.

Most recently, some researchers have introduced the DUNs to the HSI reconstruction task, which merge the advantages of both model-based and learning-based methods [13, 17, 19, 31, 32, 53]. DUNs perform the iterative process via a gradient descent module and refine the intermediate result via a deep prior module. Although existing DUNs alleviate some drawbacks of model-based and learning-based methods to some extent, there are still several bottlenecks to be solved. Firstly, how to effectively interact and fuse the features between phases is the key to the enhancement of the reconstruction quality. Most existing DUNs do not establish the connections between the proceeding and following phases. As the number of phases increases, beneficial information tends to be lost in the process of information transmission. Furthermore, inspired by the memory mechanism in the CS [28], the hidden states of the previous phases can provide complementary information for the computation of the current phase. Therefore, it is essential to introduce a feature interaction mechanism between different phases to obtain enhanced feature representation. Secondly, how to dynamically learn the parameters in the gradient descent module has been ignored in the past. Previous researchers usually treat these parameters as fixed constants to compress HSIs. However, the fixed parameters can not be adjusted adaptively and contextually in different scenes, which will lead to sub-optimal reconstruction and restrict the flexibility of DUNs. Thirdly, targeted at HSIs, the existing DUN does not combine mask optimization and image reconstruction into a united framework, which can not retain the structure and information of HSIs completely.

Inspired by the ISTA, a novel **H**yperspectral **E**xplicable **R**econstruction and **O**ptimal **S**ampling deep **Net**work for SCI, dubbed **HerosNet**, is proposed for joint mask optimization and HSI reconstruction. Particularly, the network consists of a sampling subnet, an initialization subnet, and a recovery subnet. Fueled by the success of ISTA-Net and its variant [37, 44, 49], the recovery phase updates the current estimate via a dynamic gradient descent module (DGDM) and refine the rough estimate via a hierarchical feature interaction module (HFIM). Since each recovery phase corresponds to an ISTA iteration and all the parameters are learned end-to-end, the network enjoys the merits of high-quality reconstruction with strong interpretability. For example, in Fig. 1, the results of the proposed HerosNet have clearer detailed textures than those of DGSM [13]. Overall, our contributions are summarized as follows.

- A novel ISTA inspired deep unfolding network, dubbed HerosNet, is proposed for jointly learning binary optimal masks and recovering high-quality HSIs.

- A dynamic gradient descent module (DGDM) is introduced to flexibly simulate the sensing matrix and contextually adjust the step size in gradient descent step.

- A hierarchical feature interaction module (HFIM) is designed, which fuses and interacts the hidden states of previous phases to recover the HSI frames of the current phase in proximal mapping step.

- Our HerosNet outperforms the state-of-the-art methods on simulation and real datasets by large margins.

2. Related Works

2.1. HSI Reconstruction Algorithms

Model-based Methods: The model-based methods employ the regularization term inspired by the image prior to solve the ill-posed inverse problem iteratively. In [5], a two-step iterative shrinkage/thresholding (TwIST) algorithm was designed to keep the good reconstruction performance and improve the speed of convergence. In [46], the total variation optimization was applied in the HSI reconstruction and the generalized alternating projection (GAP) algorithm was utilized to solve the optimization problem. In [15], an overcomplete dictionary was learned to exploit the sparsity representation and reconstruct the HSIs. In [40, 41], the HSI reconstruction task was treated as the maximum likelihood estimation and the Gaussian mixture model (GMM) was introduced to model the data distribution of the HSIs. Recently, in [16], the non-local self-similarity of HSIs and rank minimization strategy were incorporated into the framework of alternating direction method of multipliers (ADMM), which has achieved the best performance among the traditional model-based methods. Although these methods produce decent results in specific applications, it is difficult to design hand-crafted priors suitable for all scenes.

Deep Learning-based Methods: Relying on the powerful representation ability of deep networks, the learning-based HSI reconstruction methods have attracted more and more attention. The learning-based methods are divided into two categories generally according to whether they are unfolded from the optimization process. Among the methods without deep unfolding, the end-to-end deep networks tend to directly learn a non-linear mapping from the 2D measurement to the 3D hyperspectral cube. For instance, Miao *et*
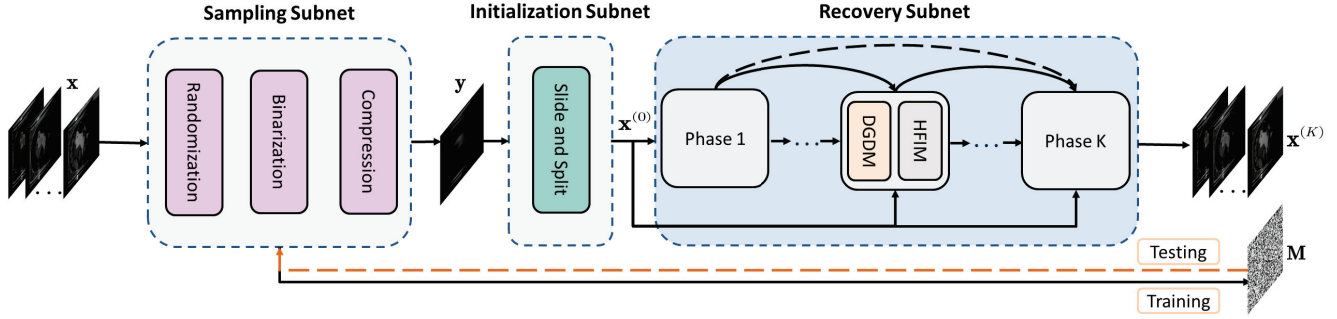
Figure 2. Illustration of the proposed HerosNet, including sampling subnet, initialization subnet and recovery subnet. The recovery subnet consists of K phases. Each phase is composed of the dynamic gradient descent module (DGDM) and the hierarchical feature interaction module (HFIM). In the training procedure, the HerosNet takes the 3D hyperspectral cube $\mathbf{x}$ as input to obtain the compressed measurement $\mathbf{y}$, and generates the reconstructed HSI frames $\mathbf{x}^{(K)}$ and the optimal mask $\mathbf{M}$. In the testing procedure, the HerosNet compresses the hyperspectral cube $\mathbf{x}$ via the optimized binary mask $\mathbf{M}$ and reconstructs HSI frames $\mathbf{x}^{(K)}$.

al. [24] introduced the dual-stage generative model to extract both spectral and spatial information. Meng *et al.* [20] embedded three self-attention modules into the U-Net backbone, thus achieving high-quality and real-time reconstruction. Apart from the end-to-end networks, the plug-and-play (PnP) framework [48, 57] incorporated the pre-trained deep denoisers into the optimization process and effectively projected the image signal to the desired domain. Most recently, deep image prior was integrated with the PnP regime to construct an untrained self-supervised network [22]. Although these methods perform a certain role in reconstructing HSIs, they all face some unavoidable challenges. For instance, the end-to-end deep networks are lack of interpretability and the PnP frameworks are very slow.

Among the deep unfolding methods, Wang *et al.* [31] unfolded the half-quadratic splitting (HQS) method and designed a spatial-spectral deep priors to boost the data fidelity. Furthermore, the local and non-local correlations of HSIs [32] were considered in the prior design. Zhang *et al.* [53] learned the tensor low-rank spectral prior via the deep CP decomposition. Most recently, Huang *et al.* [13] proposed a deep Gaussian scale mixture model to learn the scale prior and estimate the local means of images via the 3D filter. Although these methods have achieved great success, the lack of cross-phase feature interaction and content-aware parameter adjustment are still major bottlenecks for the reconstruction performance. Hereby, a novel DUN is proposed to effectively exploit the cross-phase correlation and update the parameters adaptively in this paper.

2.2. Mask Optimization Algorithms

Some existing works on traditional CS have explored the possibility of joint mask optimization and image reconstruction. For instance, Zhang *et al.* [50] proposed a constrained optimization-inspired network for adaptive sampling and recovery. You *et al.* [45] introduced the random projection augmentation strategy to learn the arbitrary-sampling matrices and improve the generalization ability of the model. In the spectral SCI, Arguello *et al.* [4] transformed mask optimization into a rank minimization problem based on the theory of the Restricted Isometry Property (RIP) [10]. Furthermore, Wang *et al.* [33] rearranged the shifted 3D data cube and divided it into four parameter-sharing sub-patches for sampling mask learning. Simultaneously, Zhang *et al.* [52] designed an end-to-end learnable auto-encoder to optimize the illumination pattern and compress the HSIs. Although the above-mentioned methods realize the adaptive sampling to a certain extent, combining mask optimization with the DUN in the spectral SCI is still challenging and worth exploring.

3. Proposed Method

3.1. Problem Formulation

In CASSI system, the 3D hyperspectral cube is firstly modulated via a coded aperture (i.e. a physical mask) and then dispersed via a dispersive prism. Mathematically, considering a sequence $\{\mathbf{X}_i\}_{i=1}^C \in \mathbb{R}^{H \times W}$ composed of C HSI frames, they are modulated via a mask $\mathbf{M} \in \mathbb{R}^{H \times W}$:

$$\mathbf{X}'_i = \mathbf{M} \odot \mathbf{X}_i, \tag{1}$$

where $\mathbf{X}'_i$ is the modulated HSI frame and $\odot$ is the Hadamard (element-wise) product. After that, modulated HSI frames with different wavelengths are shifted spatially and summed in an element-wise manner. Therefore, the modulated HSI frames $\{\mathbf{X}'_i\}_{i=1}^C \in \mathbb{R}^{H \times W}$ are compressed to a coded measurement as follows:

$$\mathbf{Y}(m, n) = \sum_{i=1}^{C} \mathbf{X}'_i(m, n + d_i) + \mathbf{N}, \tag{2}$$

where m, n denote the spatial coordinates, and d_i denotes the shifting distance of the ith channel. $\mathbf{N} \in \mathbb{R}^{H \times (W+C-1)}$
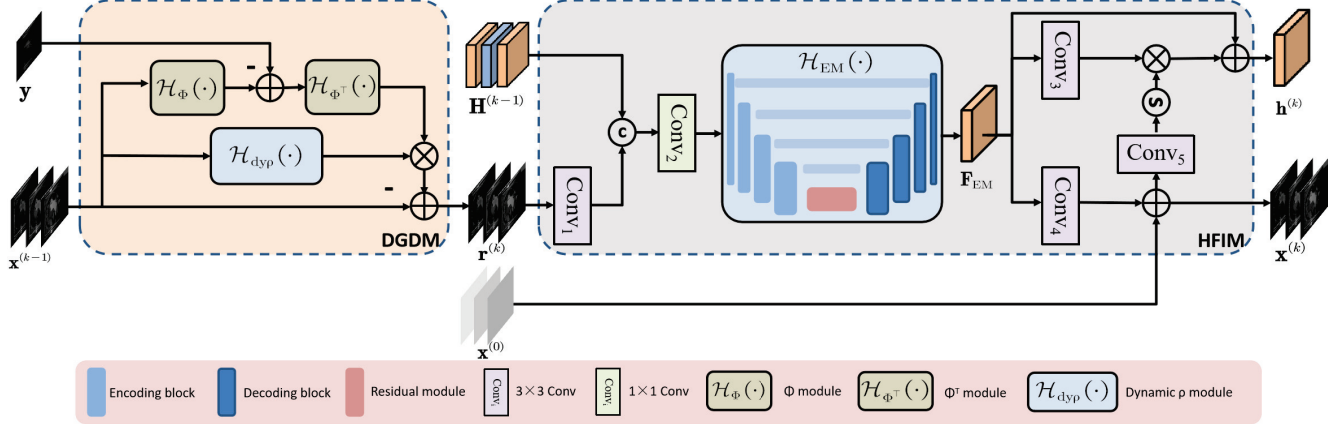
Figure 3. Details of the recovery phase in HerosNet. The recovery phase is composed of a dynamic gradient descent module (DGDM) and a hierarchical feature interaction module (HFIM). The DGDM takes the intermediate result $\mathbf{x}^{(k-1)}$ and the measurement $\mathbf{y}$ as input, and generates the coarse estimate $\mathbf{r}^{(k)}$. The HFIM is designed to refine the coarse estimate $\mathbf{r}^{(k)}$ with the hidden states of the previous phases $\mathbf{H}^{(k-1)}$ and the initialization cube $\mathbf{x}^{(0)}$ to produce the accurate reconstructed images $\mathbf{x}^{(k)}$ and the hidden states $\mathbf{h}^{(k)}$ in the kth phase.

and $\mathbf{Y} \in \mathbb{R}^{H \times (W+C-1)}$ denote the noise and the compressed measurement, respectively. The vectorized form of the spectral SCI is expressed as follows:

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{n}, \quad (3)$$

where $\mathbf{x} \in \mathbb{R}^{HWC}$, $\mathbf{y} \in \mathbb{R}^{H(W+C-1)}$, and $\mathbf{n} \in \mathbb{R}^{H(W+C-1)}$ denote the vectorized form of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{N}$, respectively. $\boldsymbol{\Phi} \in \mathbb{R}^{H(W+C-1) \times HWC}$ represents the sensing matrix.

3.2. Architecture of Proposed HerosNet

In this subsection, we propose an optimization-inspired deep unfolding network for joint mask optimization and image reconstruction. As illustrated in Fig. 2, the proposed HerosNet is composed of three subnets, including a sampling subnet, an initialization subnet, and a recovery subnet.

3.2.1 Sampling Subnet

In this paper, the sampling subnet aims to learn the optimal binary mask for the HSI compressive sensing, which preserves enough spectral-spatial information and eliminates redundancy. The training process of the sampling subnet is divided into three stages, including randomization, binarization, and compression. To learn the binary mask $\mathbf{M}$, random Gaussian initialization with the mean μ_b and the variance σ_b is adopted to generate a continuous matrix $\tilde{\mathbf{M}}$. Furthermore, we design an element-wise binarization function $\mathrm{BinarySign}(\cdot)$ to convert the continuous matrix into the binary mask as follows:

$$\mathbf{M} = \mathrm{BinarySign}(\tilde{\mathbf{M}}), \quad (4)$$

$$with \quad \mathrm{BinarySign}(z) = 1 \text{ if } z \geq \mu_b \text{ or } 0 \text{ else.} \quad (5)$$

According to the imaging rule depicted in the Sec. 3.1, we utilize a conversion function $\mathrm{Mask2Mat}(\cdot)$ to transform the binary mask $\mathbf{M}$ into the sensing matrix $\boldsymbol{\Phi}$ in Eq. (3):

$$\boldsymbol{\Phi} = \mathrm{Mask2Mat}(\mathbf{M}). \quad (6)$$

Since the sensing matrix $\boldsymbol{\Phi}$ is regarded as the learnable parameter, the derivative of the binarization function is defined as a constant, i.e. $\mathrm{BinarySign}'(z) = 1$, for the backpropagation of the sampling subnet. Finally, according to Eq. (1) and Eq. (2), the 3D hyperspectral cube $\mathbf{x}$ is compressed to the snapshot measurement $\mathbf{y}$.

3.2.2 Initialization Subnet

Given the measurement $\mathbf{y} \in \mathbb{R}^{H \times (W+C-1)}$, the initialization subnet aims to split this 2D measurement into the 3D hyperspectral cube $\mathbf{x}^{(0)} \in \mathbb{R}^{H \times W \times C}$, where W, H are the spatial size of the frames, and C is the number of spectral channels. Specifically, an extraction window is cropped from the measurement $\mathbf{y}$ and then slided in the step size of d to generate C HSI frames. Finally, the C frames are concatenated in the channel dimension to compose a 3D hyperspectral cube $\mathbf{x}^{(0)} \in \mathbb{R}^{H \times W \times C}$.

3.2.3 Recovery Subnet

As depicted in Fig. 3, the proposed recovery subnet aims to reconstruct high-quality HSIs from the compressive measurement. Inspired by the ISTA, the image reconstruction is treated as an optimization problem shown as follows:

$$\mathbf{x} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2 + \lambda\psi(\mathbf{x}). \quad (7)$$

To be noted, the first term is the data fidelity term, while the second term $\psi(\cdot)$ is the prior regularization term. λ denotes a regularization parameter.

To solve the optimization problem in Eq. (7), we unfold ISTA to design the deep network for its simplicity and interpretability. Traditional ISTA updates the results via two steps, namely gradient descent and proximal mapping, which are formulated as follows:

$$\mathbf{r}^{(k)} = \mathbf{x}^{(k-1)} - \rho\mathbf{\Phi}^{\top}(\mathbf{\Phi}\mathbf{x}^{(k-1)} - \mathbf{y}), \quad (8)$$

$$\mathbf{x}^{(k)} = \arg\min_{\mathbf{x}}\frac{1}{2}\|\mathbf{x} - \mathbf{r}^{(k)}\|_2^2 + \lambda\psi(\mathbf{x}), \quad (9)$$

where k denotes the number of ISTA iteration and ρ denotes the step size. By introducing the proximal mapping operator $\text{prox}_{\lambda\psi}(\mathbf{r}) = \arg\min_{\mathbf{x}}\frac{1}{2}\|\mathbf{x}-\mathbf{r}\|_2^2+\lambda\psi(\mathbf{x})$, Eq. (9) can be rewritten as follows:

$$\mathbf{x}^{(k)} = \text{prox}_{\lambda\psi}(\mathbf{r}^{(k)}). \quad (10)$$

Modifying these two steps, we design a dynamic gradient descent module (DGDM) and a hierarchical feature interaction module (HFIM) to reconstruct the HSIs.

Dynamic Gradient Descent Module (DGDM): To implement Eq. (8) via the deep network, the DGDM is employed to generate the immediate reconstructed results $\mathbf{r}^k$ dynamically. Most existing DUNs used to treat $\mathbf{\Phi}$, $\mathbf{\Phi}^{\top}$ and ρ in Eq. (8) as the constants, which restricts the flexibility of the network and limits the accuracy of reconstruction. To address these issues, the deep modules $\mathcal{H}_{\mathbf{\Phi}}(\cdot)$ and $\mathcal{H}_{\mathbf{\Phi}^{\top}}(\cdot)$ are introduced to simulate the matrix $\mathbf{\Phi}$ and $\mathbf{\Phi}^{\top}$ from the previous state $\mathbf{x}^{(k-1)}$, where $\mathcal{H}_{\mathbf{\Phi}}(\cdot)$ and $\mathcal{H}_{\mathbf{\Phi}^{\top}}(\cdot)$ consist of two convolution operators and four residual blocks, respectively. In order to achieve content-aware parameter adjustment, a dynamic step size operator $\mathcal{H}_{\text{dy}\rho}(\cdot)$ is incorporated into the process of gradient descent to further enhance the generalization ability of the network. The step size $\tilde{\boldsymbol{\rho}}^{(k)}$ is directly learned from the previous state $\mathbf{x}^{(k-1)}$ and adjusted adaptively with the advance of network training. Specifically, we decompose $\tilde{\boldsymbol{\rho}}^{(k)}$ into the static and dynamic components. The static component is a learnable vector and the values of each spectral channels are weight-sharing. The dynamic component is a channel attention map learned from $\mathbf{x}^{(k-1)}$. As illustrated in Fig. 4, the channel attention map is obtained by the global average pooling, two 1×1 convolution operators, ReLU activation function, and the Sigmoid function. Finally, Eq. (8) can be modified as follows:

$$\begin{aligned}\mathbf{r}^{(k)} &= \text{DGDM}(\mathbf{x}^{(k-1)}, \mathbf{y}) \\ &= \mathbf{x}^{(k-1)} - \tilde{\boldsymbol{\rho}}^{(k)}\mathcal{H}_{\mathbf{\Phi}^{\top}}(\mathcal{H}_{\mathbf{\Phi}}(\mathbf{x}^{(k-1)}) - \mathbf{y}),\end{aligned} \quad (11)$$

$$with \quad \tilde{\boldsymbol{\rho}}^{(k)} = \mathcal{H}_{\text{dy}\rho}(\mathbf{x}^{(k-1)}) = \boldsymbol{\rho}^{(k)} + \theta\boldsymbol{\Lambda}^{(k)}, \quad (12)$$
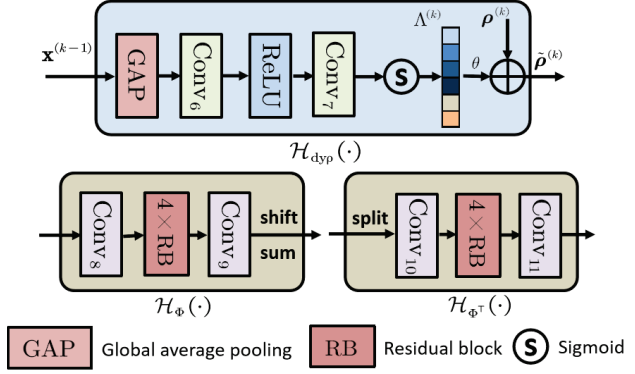

Figure 4. Details of some key components in the DGDM. The DGDM aims to simulate the sensing matrix flexibly and adjust the step size dynamically.

where $\boldsymbol{\rho}^{(k)}\in\mathbb{R}^{1\times C}$ and $\boldsymbol{\Lambda}^{(k)}\in\mathbb{R}^{1\times C}$ denote the static and the dynamic component, respectively. θ is a constant to stabilize the network training.

Hierarchical Feature Interaction Module (HFIM): To implement Eq. (9) via the deep network, the HFIM is designed to refine the coarse reconstructed result $\mathbf{r}^{(k)}$. There are two issues restricting the performance of previous deep proximal mapping modules. Firstly, since the gradient descent module functions in the image domain and the deep proximal mapping module performs in the feature domain, the spectral information will be lost when it is transmitted between these two modules. Secondly, as the number of recovery phases increases, the useful features in the previous phases are not propagated to the subsequent phases. To alleviate these issues, the proposed HFIM focuses on two aspects: 1) how to bridge the gap between the image domain and the feature domain; 2) how to effectively interact the beneficial features in previous phases to enhance the features of the current phase.

As illustrated in Fig. 3, in addition to $\mathbf{r}^{(k)}$ produced from the DGDM, the HFIM takes the cascaded hierarchical features $\mathbf{H}^{(k-1)} = [\mathbf{h}^{(k-1)}, \mathbf{h}^{(k-2)}, \ldots, \mathbf{h}^{(0)}]$ as input, where $\mathbf{h}^{(i)}$ denotes the hidden state of the ith phase. To be noted, the hidden states retain the beneficial information extracted from the reconstructed image in the current phase. Concretely, the intermediate result $\mathbf{r}^{(k)}$ is firstly transformed into the feature domain via the operator Conv_1 and fused with hidden states $\mathbf{H}^{(k-1)}$ via the dense connection Conv_2. Then, the fused feature is fed to an enhancement module $\mathcal{H}_{\text{EM}}(\cdot)$ to extract the spectral and spatial features, where $\mathcal{H}_{\text{EM}}(\cdot)$ is composed of four encoding blocks and four decoding blocks. To enhance the representation ability of the network, the residual module consisting of sixteen standard residual blocks is embedded between the encoder and decoder. The process is formulated as follows:

$$\mathbf{F}_{\text{EM}} = \mathcal{H}_{\text{EM}}(\text{Conv}_2([\mathbf{H}^{(k-1)}, \text{Conv}_1(\mathbf{r}^{(k)})])), \quad (13)$$

where $\mathbf{F}_{\text{EM}} \in \mathbb{R}^{H \times W \times N}$ denotes the enhanced feature produced by $\mathcal{H}_{\text{EM}}(\cdot)$, N denotes the channel numbers.

After obtaining the enhanced feature $\mathbf{F}_{\text{EM}} \in \mathbb{R}^{H \times W \times N}$, a well-designed feature interaction mechanism is designed to generate the reconstructed HSI frames $\mathbf{x}^{(k)} \in \mathbb{R}^{H \times W \times C}$ and the hidden state $\mathbf{h}^{(k)} \in \mathbb{R}^{H \times W \times N}$. On the one hand, the enhanced feature $\mathbf{F}_{\text{EM}}$ is transformed into the image domain via the operator Conv_4 and added with the initialization cube $\mathbf{x}^{(0)}$ to generate the reconstructed HSI frames $\mathbf{x}^{(k)}$. The process is exhibited as follows:

$$\mathbf{x}^{(k)} = \text{Conv}_4(\mathbf{F}_{\text{EM}}) + \mathbf{x}^{(0)}. \tag{14}$$

On the other hand, the network directly learns an attention cube from the current state $\mathbf{x}^{(k)}$ via the operators Conv_5 and Sigmoid to provide pixel-level interactive information and generate the hidden state $\mathbf{h}^{(k)}$, which is formulated as:

$$\mathbf{h}^{(k)} = \text{Conv}_3(\mathbf{F}_{\text{EM}}) \otimes \text{Sigmoid}(\text{Conv}_5(\mathbf{x}^{(k)})) + \mathbf{F}_{\text{EM}}. \tag{15}$$

Both $\mathbf{x}^{(k)}$ and $\mathbf{h}^{(k)}$ will be utilized in the reconstruction of subsequent phases. Finally, Eq. (9) can be modified as:

$$\mathbf{x}^{(k)}, \mathbf{h}^{(k)} = \text{HFIM}(\mathbf{r}^{(k)}, \mathbf{x}^{(0)}, \mathbf{H}^{(k-1)}). \tag{16}$$

By introducing the HFIM, we establish the hierarchical connection to integrate features in different phases. With the proposed DGDM and HFIM, the recovery subnet can reconstruct HSIs accurately and rapidly.

3.3. Network Training and Implementation Details

In our implementation, the number of spectral channels C is 28 and the number of feature channels N is 32. In our sampling subnet with mask optimization, μ_b and σ_b are set to 0 and 0.1, respectively. In our recovery subnet, the number of recovery phases K is 8 and θ in Eq. (12) is 0.5. The learnable parameters in the proposed network are denoted by Θ, including the binary mask $\mathbf{M}$, the parameters $\mathbb{D}^{(k)}$ in the DGDM and $\mathbb{H}^{(k)}$ in the HFIM. To learn the parameters $\Theta = \{\mathbf{M}, \mathbb{D}^{(k)}, \mathbb{H}^{(k)}\}$, we utilize the reconstructed results of the final phase and some intermediate phases to calculate the loss function [19]. Specifically, given the training data $\{\mathbf{x}_i\}_{i=1}^{N_d}$, the loss function is defined:

$$\mathcal{L} = \mathcal{L}_f + \mathcal{L}_p, \tag{17}$$

$$with \quad \mathcal{L}_f = \frac{1}{N_d} \sum_{i=1}^{N_d} \|\mathbf{x}_i^{(K)} - \mathbf{x}_i\|_2^2, \tag{18}$$

$$\mathcal{L}_p = \frac{\beta}{N_d} \sum_{i=1}^{N_d} \|\mathbf{x}_i^{(K-1)} - \mathbf{x}_i\|_2^2 + \|\mathbf{x}_i^{(K-2)} - \mathbf{x}_i\|_2^2. \tag{19}$$

where $\mathcal{L}_f$ and $\mathcal{L}_p$ denote the loss function of the final phase and the previous few phases, respectively. K and N_d denote
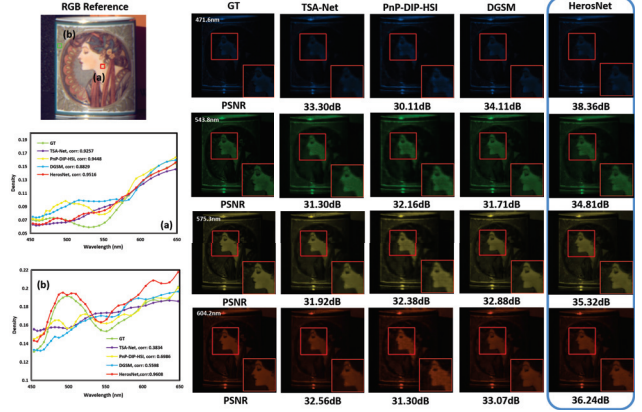

Figure 5. Reconstructed images of *Scene1* on KAIST dataset by our HerosNet and other state-of-the-art methods. The HSI frames are converted to RGB images via the CIE color matching function [27]. Spectral curves on the selected regions ((a) and (b)) and the visualization results show that our results have higher spectral accuracy and better perceptual quality.

the number of recovery phases and training samples. β is the equilibrium constant and set to 0.5.

Our HerosNet is implemented with 4 NVIDIA Tesla V100 GPUs. We adopt Adam [9] to train the network for 100 epochs. The learning rate is initialized with 1×10^{-4} and decays with a factor of 0.9 every 10 epochs.

4. Experimental Results

4.1. Experimental Settings

In this paper, we have verified the effectiveness of the proposed network on both simulation datasets and the real dataset. Following the settings of TSA-Net [20] and DGSM [13], the simulation experiments are conducted on the public HSI datasets CAVE [42] and KAIST [7] with the size $256 \times 256 \times 28$, i.e., 28 spectral channels with the spatial size 256×256. For the experiments in the real scenes, 5 compressive measurements with the spatial size of 640×694 captured by the real SCI system are utilized for testing. The metrics of PSNR and SSIM [35] are employed to evaluate the reconstruction quality.

4.2. Simulation Results

To demonstrate the effectiveness of the proposed method on the simulation datasets, we compare the proposed Heros-Net with several existing methods, including the model-based methods (GAP-TV [46], DeSCI [16]), the end-to-end deep learning-based methods (λ-net [24] and TSA-Net [20]), the plug-and-play method (PnP-DIP-HSI [22]) and the deep unfolding networks (HSSP [31] and DGSM [13]). All of these competing methods are trained on the CAVE dataset with a fixed real mask [13].

Table 1. Comparison results of the proposed network and state-of-the-art HSI reconstruction methods on the KAIST dataset. To be noted, *HerosNet-base* denotes the proposed method without mask optimization. Best results are in red and the second-best ones are in blue.

| Testing Set	GAP-TV [46] (ICIP, 2016)	DeSCI [16] (TPAMI, 2018)	HSSP [31] (CVPR, 2019)	λ-net [24] (ICCV, 2019)	TSA-Net [20] (ECCV, 2020)	PnP-DIP-HSI [22] (ICCV, 2021)	DGSM [13] (CVPR, 2021)	HerosNet-base (Ours)	HerosNet (Ours)
Scene01	25.13 / 0.724	27.15 / 0.794	31.07 / 0.852	30.82 / 0.880	31.26 / 0.887	32.70 / 0.898	33.17 / 0.954	34.24 / 0.963	**35.69 / 0.973**
Scene02	20.67 / 0.630	22.26 / 0.694	26.30 / 0.798	26.30 / 0.846	26.88 / 0.855	27.27 / 0.832	31.61 / 0.933	32.94 / 0.952	**35.01 / 0.968**
Scene03	23.19 / 0.757	26.56 / 0.877	29.00 / 0.875	29.42 / 0.916	30.03 / 0.921	31.32 / 0.920	31.55 / 0.952	34.15 / 0.966	**34.82 / 0.967**
Scene04	35.13 / 0.870	39.00 / 0.965	38.24 / 0.926	37.37 / 0.962	39.90 / 0.964	**40.79 / 0.970**	37.43 / 0.981	38.80 / 0.984	38.07 / **0.985**
Scene05	22.31 / 0.674	24.80 / 0.778	27.98 / 0.827	27.84 / 0.866	28.89 / 0.878	29.81 / 0.903	29.43 / 0.927	31.39 / 0.953	**33.18 / 0.969**
Scene06	22.90 / 0.635	23.55 / 0.753	29.16 / 0.823	30.69 / 0.886	31.30 / 0.895	30.41 / 0.890	32.49 / 0.960	32.88 / 0.960	**34.94 / 0.976**
Scene07	17.98 / 0.670	20.03 / 0.772	24.11 / 0.851	24.20 / 0.875	25.16 / 0.887	28.18 / 0.913	30.64 / 0.937	32.79 / 0.963	**33.58** / 0.962
Scene08	23.00 / 0.624	20.29 / 0.740	27.94 / 0.831	28.86 / 0.880	29.69 / 0.887	29.45 / 0.885	31.06 / 0.955	31.11 / 0.953	**33.19 / 0.968**
Scene09	23.36 / 0.717	23.98 / 0.818	29.14 / 0.822	29.32 / 0.902	30.03 / 0.903	**34.55** / 0.932	30.87 / 0.951	31.58 / 0.953	33.04 / **0.964**
Scene10	23.70 / 0.551	25.94 / 0.666	26.44 / 0.740	27.66 / 0.843	28.32 / 0.848	28.52 / 0.863	31.34 / 0.955	31.64 / 0.949	**33.01 / 0.965**
Average	23.73 / 0.683	25.86 / 0.785	28.93 / 0.834	29.25 / 0.886	30.15 / 0.893	31.30 / 0.901	31.96 / 0.951	33.15 / 0.960	**34.45 / 0.970**
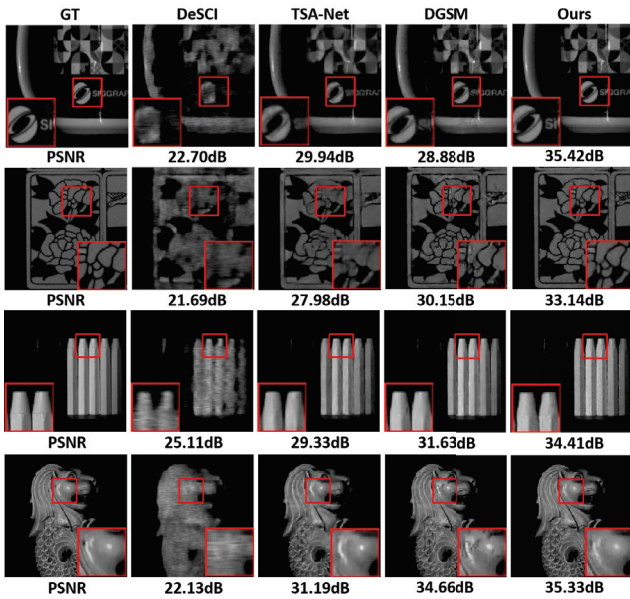
Figure 6. Visual comparisons of our HerosNet and other state-of-the-art methods on the KAIST dataset. The wavelengths of all these images are 648.1nm. The reconstructed images of HerosNet can preserve more details and clearer textures with less artifacts.

As shown in Table 1, the proposed HerosNet obtains 34.45 dB of PSNR and 0.970 of SSIM, which has surpassed all of competing methods by large margins. Compared with the state-of-the-art method DGSM, the proposed network achieves 2.49dB improvement of PSNR and 0.019 improvement of SSIM. In comparison with the second best method PnP-DIP-HSI, the proposed method also achieves 3.15dB/0.069 gains on PSNR/SSIM. It demonstrates that the proposed recovery subnet can better unearth the spectral information and the proposed mask optimization strategy can search for the optimal binary mask. As illustrated in Fig. 5, the reconstructed HSIs produced by the Heros-Net have more spatial details and clearer texture in different spectral channels, while the results of other competing

Table 2. Comparison results of the proposed optimized mask and other kinds of fixed masks. To be noted, even training on the same real mask [13], the proposed method has surpassed all of the existing methods listed in Table 1.

Mask Type	PSNR	SSIM
Uniform mask	31.78	0.935
Guassian mask	32.49	0.943
Real mask [13]	33.15	0.960
Optimized binary mask	**34.45**	**0.970**

Figure 7. Visual illustrations of different kinds of masks, including the uniform mask, Gaussian mask, real fixed mask, optimized binary mask (Ours).

methods are blurry. In addition, the spectral curves of the HerosNet have a higher correlation with the reference spectra. Fig. 6 further plots some visual comparisons of the proposed HerosNet, DGSM, TSA-Net and DeSCI on four other scenes. Compared with these three typical methods, the proposed HerosNet provides sharper edges, better visual effects and less artifacts.

4.3. Multiple Mask Results

To objectively evaluate the effectiveness of the proposed mask optimization strategy, we train our model on different kinds of masks. As shown in Fig. 7, the center areas of four different masks with the spatial size 64×64 in the experiments are presented. To be noted, we remove the mask optimization strategy from the proposed network when training on the fixed mask. Table 2 lists the PSNR and SSIM results by testing on the KAIST dataset. It can be clearly seen
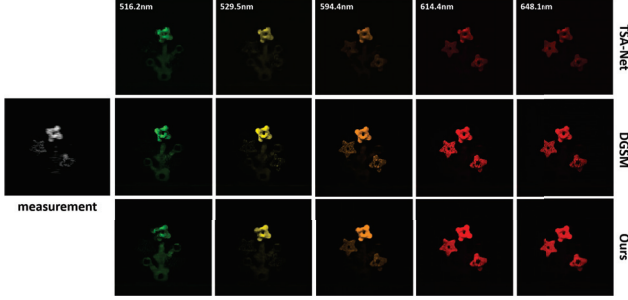
Figure 8. Visual comparisons of our HerosNet, DGSM [13] and TSA-Net [20] on the real dataset *Scene1*. Obviously, our HerosNet can recover more image details and clearer content (i.e. the flower in the right corner) than DGSM and TSA-Net.

that the proposed HerosNet with joint mask optimization outperforms the networks that are directly trained on the fixed masks. The main reason is that the optimized binary mask preserves the complete image structure and sufficient detailed information to achieve optimal sampling. Simultaneously, the proposed method trained on the real mask also surpasses all the SOTA methods listed in Table 1. It further proves that even training on the same mask, the proposed method also has great advantages in HSI reconstruction.

4.4. Real Data Results

To verify the effect of the proposed method on the real data, five compressive measurements captured by the real spectral SCI system are utilized for testing. Due to the ground truths of real scenes are unavailable, we get rid of the sampling subnet and only take the 2D compressive measurements as input. For fair comparisons, all of the methods are trained on the CAVE datasets using the fixed real mask with 11-bit shot noise injected. Fig. 8 plots the visual comparisons of the proposed HerosNet and the existing SOTA method DGSM [13], TSA-Net [20]. Obviously, our results recover more image details with fewer artifacts and clearer content in various wavelengths, while the reconstructed images of other methods are blurrier and more susceptible to noise corruption. It further proves that the proposed HerosNet is more robust to the noise distortion and effective in the real spectral imaging system.

4.5. Ablation study

To evaluate the contribution of different components in the proposed HerosNet, ablation study is conducted on the CAVE and KAIST datasets. We mainly focus on the four components, namely, mask optimization (MO), hierarchical feature interaction module (HFIM), residual module (RM), and dynamic step size (Dyρ) mechanism. Table 3 shows the results of PSNR and SSIM on different settings. To investigate the validity of mask optimization strategy, we remove the MO and retrain our reconstruction network on

Table 3. Evaluation of the effectiveness of different components.

Case Index	MO	HFIM	RM	Dyρ	PSNR	SSIM
(a) (Ours-base)	×	✓	✓	✓	33.15	0.960
(b)	✓	×	✓	✓	33.41	0.961
(c)	✓	✓	×	✓	33.59	0.965
(d)	✓	✓	✓	×	34.11	0.966
(e) (Ours)	✓	✓	✓	✓	34.45	0.970

the fixed real mask. It can be clearly seen that the PSNR and SSIM results have a decline by 1.30dB and 0.010 respectively, which proves the effectiveness of the proposed MO. To be noted, even without the MO, *Ours-base* also achieves the best performance among all the existing reconstruction methods listed in Table 1. To investigate the impact of HFIM, we reimplement a variant network, which directly utilizes a U-Net as the deep prior without any interaction between phases. Obviously, without HFIM, the values of PSNR and SSIM have dropped by 1.04dB and 0.009 respectively, thus proving its significant effect. Meanwhile, Table 3 shows a substantial drop on PSNR/SSIM from 34.45dB/0.970 to 33.59dB/0.965 when RM is removed. Furthermore, replacing the dynamic step size $\tilde{\rho}^{(k)}$ with the static component $\rho^{(k)}$ in Eq. (12), the results of PSNR and SSIM have decreased by 0.34dB and 0.004 respectively, which verifies the role of dynamic step size mechanism.

5. Conclusion

In this paper, we propose a novel HerosNet for spectral snapshot compressive imaging. Inspired by the ISTA, HerosNet unfolds the optimization iterative process and is capable of jointly optimizing binary masks and reconstructing the HSIs accurately. To improve the generalization ability and flexibility of the network, a dynamic gradient descent module is proposed to achieve adaptive and content-aware parameter adjustment. To better utilize the cross-phase correlation, a hierarchical feature interaction module is designed to fuse and interact the useful information between different phases. Finally, experiments demonstrate that our network outperforms the state-of-the-art methods on both simulation and real datasets. Our future work will support HerosNet on MindSpore [1], which is a new deep learning computing framework.

Broader impacts and limitations: The proposed HerosNet contributes to the industrial application of spectral SCI and inspires the design of deep unfolding networks in other image inverse problems. Whereas, our model can not obtain decent results without retraining or fine-tuning when it comes to different imaging systems and physical masks. Meanwhile, the proposed learning-based method will inevitably reflect biases in the training data. These issues warrant further research and exploration for application.

References

[1] Mindspore. https://www.mindspore.cn/, 2020.

[2] Hamed Akbari, Yukio Kosugi, Kazuyuki Kojima, and Naofumi Tanaka. Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging. *IEEE Transactions on Biomedical Engineering*, 57(8):2011–2017, 2010.

[3] Gonzalo R Arce, David J Brady, Lawrence Carin, Henry Arguello, and David S Kittle. Compressive coded aperture spectral imaging: an introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2013.

[4] Henry Arguello and Gonzalo R Arce. Rank minimization code aperture design for spectrally selective compressive imaging. *IEEE Transactions on Image Processing*, 22(3):941–954, 2012.

[5] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, 2007.

[6] Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan, Qionghai Dai, Lawrence Carin, and David J Brady. Computational snapshot multispectral cameras: toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108, 2016.

[7] Inchang Choi, MH Kim, D Gutierrez, DS Jeon, and G Nam. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics*, 36(6):1–13, 2017.

[8] Meng Ding, Xiao Fu, Ting-Zhu Huang, Jun Wang, and Xi-Le Zhao. Hyperspectral super-resolution via interpretable block-term tensor modeling. *IEEE Journal of Selected Topics in Signal Processing*, 15(3):641–656, 2020.

[9] Kingma DP and Jimmy Ba. Adam: a method for stochastic optimization. In *Proceedings of the International Conference for Learning Representations*, 2015.

[10] Yonina C Eldar and Gitta Kutyniok. *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[11] Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express*, 15(21):14013–14027, 2007.

[12] Wei He, Naoto Yokoya, and Xin Yuan. Fast hyperspectral image recovery of dual-camera compressive hyperspectral imaging via non-iterative subspace-based fusion. *IEEE Transactions on Image Processing*, 30:7170–7183, 2021.

[13] Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi. Deep gaussian scale mixture prior for spectral compressive imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[14] Jie Lei, Weiying Xie, Jian Yang, Yunsong Li, and Chein-I Chang. Spectral-spatial feature extraction for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10):8131–8143, 2019.

[15] Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics*, 33(6):1–11, 2014.

[16] Yang Liu, Xin Yuan, Jinli Suo, David J Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2990–3006, 2018.

[17] Jiawei Ma, Xiao-Yang Liu, Zheng Shou, and Xin Yuan. Deep tensor admm-net for snapshot compressive imaging. In *Proceedings of the IEEE International Conference on Computer Vision*, 2019.

[18] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, 2004.

[19] Ziyi Meng, Shirin Jalali, and Xin Yuan. Gap-net for snapshot compressive imaging. *arXiv preprint arXiv:2012.08364*, 2020.

[20] Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *Proceedings of European Conference on Computer Vision*, 2020.

[21] Ziyi Meng, Mu Qiao, Jiawei Ma, Zhenming Yu, Kun Xu, and Xin Yuan. Snapshot multispectral endomicroscopy. *Optics Letters*, 45(14):3897–3900, 2020.

[22] Ziyi Meng, Zhenming Yu, Kun Xu, and Xin Yuan. Self-supervised neural networks for spectral snapshot compressive imaging. In *Proceedings of the IEEE International Conference on Computer Vision*, 2021.

[23] Ziyi Meng and Xin Yuan. Perception inspired deep neural networks for spectral snapshot compressive imaging. In *Proceedings of the IEEE International Conference on Image Processing*, 2021.

[24] Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos. λ-net: reconstruct hyperspectral images from a snapshot measurement. In *Proceedings of the IEEE International Conference on Computer Vision*, 2019.

[25] Lujendra Ojha, Mary Beth Wilhelm, Scott L Murchie, Alfred S McEwen, James J Wray, Jennifer Hanley, Marion Massé, and Matt Chojnacki. Spectral evidence for hydrated salts in recurring slope lineae on mars. *Nature Geoscience*, 8(11):829–832, 2015.

[26] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018.

[27] Thomas Smith and John Guild. The cie colorimetric standards and their use. *Transactions of the Optical Society*, 33(3):73, 1931.

[28] Jiechong Song, Bin Chen, and Jian Zhang. Memory-augmented deep unfolding network for compressive sensing. In *Proceedings of the ACM International Conference on Multimedia*, 2021.

[29] Jin Tan, Yanting Ma, Hoover Rueda, Dror Baron, and Gonzalo R Arce. Compressive hyperspectral imaging via approximate message passing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):389–401, 2015.

[30] Ashwin A Wagadarikar, Nikos P Pitsianis, Xiaobai Sun, and David J Brady. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics Express*, 17(8):6368–6388, 2009.

[31] Lizhi Wang, Chen Sun, Ying Fu, Min H Kim, and Hua Huang. Hyperspectral image reconstruction using a deep spatial-spectral prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[32] Lizhi Wang, Chen Sun, Maoqing Zhang, Ying Fu, and Hua Huang. Dnu: deep non-local unrolling for computational spectral imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[33] Lizhi Wang, Tao Zhang, Ying Fu, and Hua Huang. Hyper-reconnet: joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Transactions on Image Processing*, 28(5):2257–2270, 2018.

[34] Minghua Wang, Qiang Wang, and Jocelyn Chanussot. Tensor low-rank constraint and l_0 total variation for hyperspectral image mixed noise removal. *IEEE Journal of Selected Topics in Signal Processing*, 15(3):718–733, 2021.

[35] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[36] Xueling Wei, Wei Li, Mengmeng Zhang, and Qingli Li. Medical hyperspectral image classification based on end-to-end fusion deep neural network. *IEEE Transactions on Instrumentation and Measurement*, 68(11):4481–4492, 2019.

[37] Zhuoyuan Wu, Jian Zhang, and Chong Mou. Dense deep unfolding network with 3d-cnn prior for snapshot compressive imaging. In *Proceedings of the IEEE International Conference on Computer Vision*, 2021.

[38] Weiying Xie, Tao Jiang, Yunsong Li, Xiuping Jia, and Jie Lei. Structure tensor and guided filtering-based algorithm for hyperspectral anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):4218–4230, 2019.

[39] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017.

[40] Jianbo Yang, Xuejun Liao, Xin Yuan, Patrick Llull, David J Brady, Guillermo Sapiro, and Lawrence Carin. Compressive sensing by learning a gaussian mixture model from measurements. *IEEE Transactions on Image Processing*, 24(1):106–119, 2014.

[41] Jianbo Yang, Xin Yuan, Xuejun Liao, Patrick Llull, David J Brady, Guillermo Sapiro, and Lawrence Carin. Video compressive sensing using gaussian mixture models. *IEEE Transactions on Image Processing*, 23(11):4863–4878, 2014.

[42] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253, 2010.

[43] Kouhei Yorimoto and Xian-Hua Han. Hypermixnet: hyperspectral image reconstruction with deep mixed network from a snapshot measurement. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2021.

[44] Di You, Jingfen Xie, and Jian Zhang. Ista-net++: flexible deep unfolding network for compressive sensing. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2021.

[45] Di You, Jian Zhang, Jingfen Xie, Bin Chen, and Siwei Ma. Coast: controllable arbitrary-sampling network for compressive sensing. *IEEE Transactions on Image Processing*, 30:6066–6080, 2021.

[46] Xin Yuan. Generalized alternating projection based total variation minimization for compressive sensing. In *Proceedings of IEEE International Conference on Image Processing*, 2016.

[47] Xin Yuan, David J Brady, and Aggelos K Katsaggelos. Snapshot compressive imaging: theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 38(2):65–88, 2021.

[48] Xin Yuan, Yang Liu, Jinli Suo, and Qionghai Dai. Plug-and-play algorithms for large-scale snapshot compressive imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[49] Jian Zhang and Bernard Ghanem. Ista-net: interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[50] Jian Zhang, Chen Zhao, and Wen Gao. Optimization-inspired compact deep compressive sensing. *IEEE Journal of Selected Topics in Signal Processing*, 14(4):765–774, 2020.

[51] Jian Zhang, Chen Zhao, Debin Zhao, and Wen Gao. Image compressive sensing recovery using adaptively learned sparsifying basis via l_0 minimization. *Signal Processing*, 103:114–126, 2014.

[52] Kangning Zhang, Junjie Hu, and Weijian Yang. Deep compressed imaging via optimized pattern scanning. *Photonics Research*, 9(3):B57–B70, 2021.

[53] Shipeng Zhang, Lizhi Wang, Lei Zhang, and Hua Huang. Learning tensor low-rank prior for hyperspectral image reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[54] Chen Zhao, Siwei Ma, Jian Zhang, Ruiqin Xiong, and Wen Gao. Video compressive sensing reconstruction via reweighted residual sparsity. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(6):1182–1195, 2016.

[55] Ji Zhao, Yanfei Zhong, Yunyun Wu, Liangpei Zhang, and Hong Shu. Sub-pixel mapping based on conditional random fields for hyperspectral remote sensing imagery. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1049–1060, 2015.

[56] Min Zhao, Longbin Yan, and Jie Chen. Lstm-dnn based autoencoder network for nonlinear hyperspectral image unmixing. *IEEE Journal of Selected Topics in Signal Processing*, 15(2):295–309, 2021.

[57] Siming Zheng, Yang Liu, Ziyi Meng, Mu Qiao, Zhishen Tong, Xiaoyu Yang, Shensheng Han, and Xin Yuan. Deep plug-and-play priors for spectral snapshot compressive imaging. *Photonics Research*, 9(2):B18–B29, 2021.